

High quality, sensitive detection of copy number variations (CNVs) at high throughput on Agilent microarrays



Graham Speight, Nicole Sparkes, Andrew J. Rogers, Sandra G.S. Lam,
Tom J. Nicholls, John Shovelton, Brian J. Eddington, John Anson

Introduction

Differences in genomic copy numbers have been known and studied for many years^{1,2}, with increasing evidence that small chromosomal imbalances leading to genomic copy number variation (CNV) are common phenomena, affecting as much as 12 % of the human genome³. In addition to reports of the presence of CNVs in healthy individuals⁴, many studies have associated CNVs with disease phenotypes, particularly for neurobiological conditions such as autism⁵ and schizophrenia⁶.

Large association studies or family (trio) based genetic analyses are required to elucidate the genetic components for these complex diseases, often requiring thousands of patient and control samples. Traditional genotyping methods, such as karyotyping and fluorescent *in situ* hybridisation (FISH), do not offer the resolution or throughput required for these investigations, creating a need for new techniques permitting highly accurate, reproducible and sensitive screening of CNVs in a high throughput format.

Several novel technologies now exist which enable genome-wide association studies (GWAS) for CNV analysis, with oligonucleotide microarray-based comparative genomic hybridisation (aCGH) now considered the gold standard technology for this application⁷. One of the major advantages of aCGH is the potential to custom design probes for a specific genomic region, offering virtually single base resolution.

Methods

Until recently, aCGH has only been used for CNV analysis of small numbers of samples, as scale-up of these manual techniques is technically challenging, time consuming and has the potential for high levels of experimental error. The major restrictions of these techniques include:

- absence of accurate and reproducible digestion, labelling and hybridisation protocols for liquid handling robotics;
- dependency on single column purification for unincorporated dye removal;
- difficulty in tracking all samples, reagent batches, arrays and equipment;
- need for suitable ozone monitoring and control;
- absence of reproducible multiple slide washing protocols for automated systems;
- large computational capability required for analysing, storing and creating back-ups of large quantities of raw and processed data.

Oxford Gene Technology (OGT), in collaboration with Agilent Technologies, has developed an aCGH protocol to permit high throughput analysis. OGT's high throughput microarray facility is capable of analysing in excess of 1,250 samples and references a week, with a modular design to allow future scale-up of capacity as required. The high throughput microarray service protocol incorporates:

- automated pipetting protocols to replace all manual pipetting steps, minimise experimental error and ensure optimal reproducibility;
- 96-well format batch processing dye removal protocol, increasing reproducibility and minimising the risk of sample mix-up;
- a proprietary, flexible laboratory information management system (LIMS) to track samples, reagents and equipment throughout the procedure that also includes automated analysis and QC assessment, as well as built-in validation steps. Spurious results are immediately highlighted, allowing prompt corrective measures to safeguard precious samples;
- an ozone-monitored and controlled environment (< 5 ppb) to prevent fluorescent signal degradation, reducing the chance of variability and potential data loss;
- automated washing workstations, using specific batch processing protocols for Agilent aCGH slides, reducing variability and increasing reproducibility;



- immediate back-up of TIFF images for all processed arrays, with real-time feature extraction, ensuring data is securely stored at multiple on- and off-site locations;
- electrical supply protection, with a back-up generator on site, minimising delays and reducing the risk to precious samples;
- GLP-like working practices, with 20 LIMS-tracked sample QC steps, > 30 in-process QC steps and 15 manager sign off/validation steps.

Results

OGT recently used its high throughput technology to process in excess of 22,000 samples under very demanding timelines, generating almost 3 billion data points to exacting quality requirements.

A majority of the samples were part of the Wellcome Trust Case Control Consortium (WTCCC) CNV study⁸. This is the world's largest CNV study – involving 24 leading human geneticists – using DNA samples from over 21,000 patients and controls to identify genetic variants influencing susceptibility to diseases, including tuberculosis, coronary heart disease, types 1 and 2 diabetes, rheumatoid arthritis, Crohn's disease, bipolar disorder, autoimmune thyroid disease, ankylosing spondylitis, multiple sclerosis, breast cancer and hypertension.

An internal control was run on every plate to continually monitor the quality and performance of the high throughput processing, with over 40 quality control checks performed and documented for each sample, providing evidence of the excellent QC metrics of the workflow (see table 1 and figures 1-5).

Table 1 Agilent array QC metrics for the internal control from over 400 plates of aCGH data.

This demonstrates the excellent QC metrics of OGT's high throughput protocols and workflow across the broad spectrum of array quality metrics monitored over a large number of samples.

Metric measured	% passed excellence criteria
Derivative log ratio spread	98
Background noise green	99
Background noise red	99
Signal intensity green	100
Signal intensity red	100
Signal-to-noise green	99
Signal-to-noise red	99

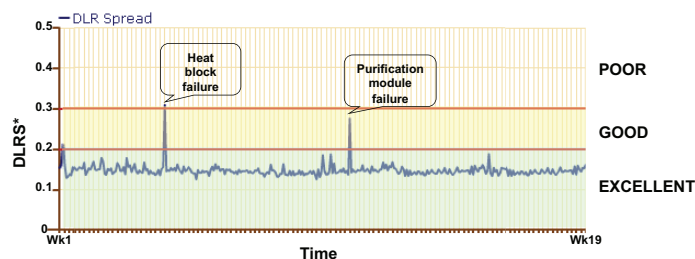


Figure 1 Derivative log ratio spread (DLRS) of control samples run on more than 400 plates.

This metric is a measure of the reliability of the data to detect and call CNV aberrations. Data illustrated is from 19 weeks of a high throughput CNV project showing a 98 % pass rate of control samples run on every plate. Where there is deviation from excellent data, the causative piece of equipment or consumable can be rapidly identified using OGT's bespoke LIMS and removed from the processing, ensuring an immediate return to high quality data.

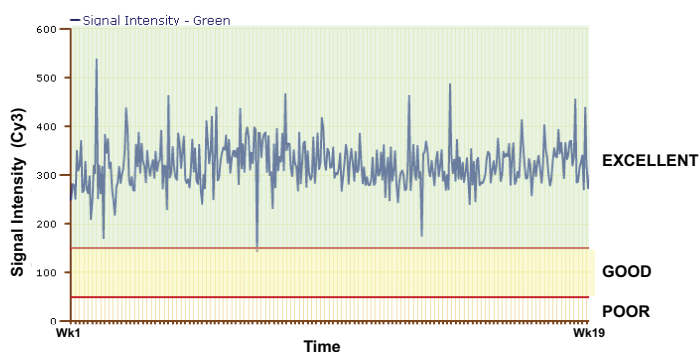


Figure 2 Signal intensity of control samples over a period of 19 weeks where more than 400 plates were processed.

Signal intensity is consistently high and gives excellent QC metrics (based on Agilent QC metrics) throughout the 19 weeks of processing through OGT's high throughput protocols and workflow. This allows consistent, reproducible generation of data throughout large scale projects.

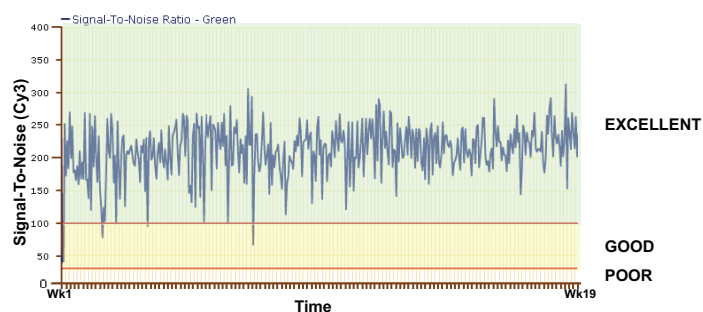


Figure 3 Signal-to-noise ratio of control samples.

Signal-to-noise is consistently excellent (based on Agilent QC metrics) throughout 19 weeks (> 400 plates) of processing. This ensures optimal dynamic range of data, enabling greater confidence in CNV calling throughout large scale projects.

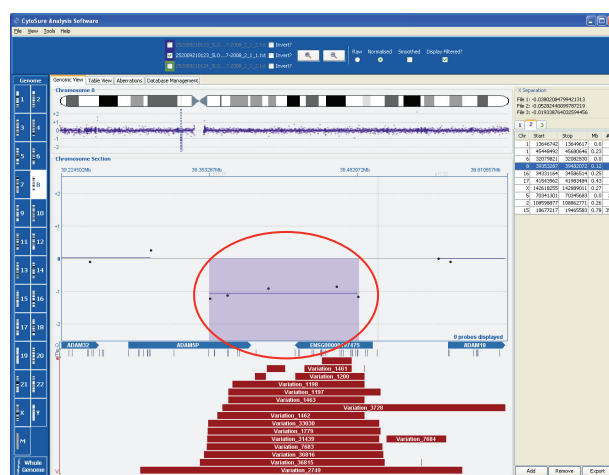


Figure 4 Screenshot of the CytoSure™ Interpret software.

Indicating a ~ 1 Mb deletion in a known CNV region on chromosome 15. Multiple samples can be loaded into the database to allow population distribution analysis of CNVs.

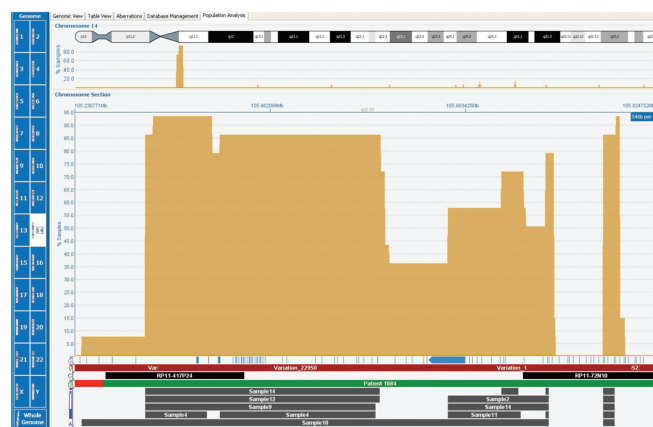


Figure 5 Screenshot of the CytoSure Interpret software indicating a CNV detected in a number of different samples (population).

Multiple samples can be loaded into the database and viewed simultaneously. This allows the population distribution of specific CNVs to be displayed. The length and localisation of the identified CNV can be shown for each sample (as displayed in the lower sample track in the software).

Conclusions

OGT processed over 20,000 samples from the landmark WTCCC CNV study in 20 weeks, using state-of-the-art automated processing to achieve exceptional data quality from custom whole-genome human CNV-focussed microarrays (developed by Agilent) throughout the entirety of the project.

OGT is committed to providing high quality data for a variety of high throughput microarray applications, including miRNA, methylation, gene expression and aCGH, and many scientists have benefited from OGT's bespoke microarray service, from array design and synthesis through to bioinformatics support and data analysis.

References

- 1 Ottolenghi, S., Lanyon, W.G., Paul, J., Williamson, R., Weatherall, D.J., Clegg, J.B., Pritchard, J., Pootrakul, S., and Boon, W.H. (1974). The severe form of alpha thalassaemia is caused by a haemoglobin gene deletion. *Nature* 251, 389–392.
- 2 Cairns, P., Polascik, T.J., Eby, Y., Tokino, K., Califano, J., Merlo, A., Mao, L., Herath, J., Jenkins, R., Westra, W., Rutter, J.L., Buckler, A., Gabrielson, E., Tockman, M., Cho, K.R., Hedrick, L., Bova, G.S., Isaacs, W., Koch, W., Schwab, D., Sidransky, D. (1995). Frequency of homozygous deletion at p16/CDKN2 in primary human tumours. *Nature Genetics* 11, 210 – 212.
- 3 Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurler, M.E. (2006). Global variation in copy number in the human genome. *Nature* 444, 444-454.
- 4 Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M., Frackelton, E.C., Geiger, E.A., Haldeman-Englert, C., Imielinski, M., Kim, C.E., Medne, L., Annaiah, K., Bradford, J.P., Dabaghyan, E., Eckert, A., Onyiah, C.C., Ostapenko, S., Otieno, F., Santa, E., Shaner, J.L., Skraban, R., Smith, R.M., Elia, J., Goldmuntz, E., Spinner, N.B., Zackai, E.H., Chiavacci, R.M., Grundmeier, R., Rappaport, E.F., Grant, S.F.A., White, P.S., Hakonarson, H. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Research* 19, 1682-1690.
- 5 Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimäki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K., Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science* 316, 445-449.
- 6 Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., International Schizophrenia Consortium, the Wellcome Trust Case Control Consortium, Craddock, N., Owen, M.J., O'Donovan, M.C. (2009). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Human Molecular Genetics* 18, 1497-1503.
- 7 Carter, N.P. (2007). Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature Genetics* 39 (7 suppl), S16-S21.
- 8 Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., MacArthur, D.G., MacDonald, J.R., Onyiah, I., Pang, A.W.C., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., The Wellcome Trust Case Control Consortium, Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., Hurler, M.E. (2009). Origins and functional impact of copy number variation in the human genome. *Nature advance online publication* Oct 7, doi:10.1038/nature08516.



Genomic Services from OGT: Combining industry leading platforms, expert people and unparalleled process power – to rapidly deliver high quality genomic data to you.
We call this Genefficiency™

Oxford Gene Technology T: +44 (0)1865 856826 E: genomicservices@ogt.co.uk W: www.ogt.co.uk

This document and its contents are © Oxford Gene Technology IP Limited - 2009. All rights reserved. OGT™, CytoSure™ and Oxford Gene Technology are trademarks of Oxford Gene Technology IP Limited.